

Rec'd PCT/PTO 23 MAR 2005

10/528675

Aldehyde dehydrogenase gene

The present invention relates to a novel DNA which encodes aldehyde dehydrogenase (SNDH) derived from *Gluconobacter oxydans* DSM 4025, an expression vector containing the DNA and a recombinant microorganism containing the expression vector. Furthermore, the present invention concerns a process for producing
5 recombinant aldehyde dehydrogenase protein and a process for producing L-ascorbic acid (vitamin C) and/or 2-keto-L-gulonic acid (2-KGA) from L-sorbose by using the recombinant aldehyde dehydrogenase protein or the recombinant microorganism containing said expression vector.

Vitamin C is one of indispensable nutrient factors for human beings and has been
10 commercially synthesized by the Reichstein process for about 60 years. Synthetic vitamin C is also used in animal feed even though farm animals can synthesize it in their own body. Although the Reichstein process has many advantageous points for industrial vitamin C production, it still has undesirable problems such as high energy consumption and usage of considerable quantities of organic and inorganic solvents. Therefore, over
15 the past decades, many approaches to manufacture vitamin C using enzymatic conversions, which would be more economical as well as ecological, have been investigated.

The present invention is directed to an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to
20 the nucleotide sequence of SEQ ID NO: 1.

As used herein, "SNDH" stands for aldehyde dehydrogenase.

As used herein, "nucleic acid molecule" includes both DNA and RNA and, unless otherwise specified, includes both double-stranded, single-stranded nucleic acid, and nucleosides thereof. Also included are hybrids such as DNA-RNA hybrids, DNA-RNA-
25 protein hybrids, RNA-protein hybrids, and DNA-protein hybrids.

As used herein, "mutation" refers to a single base pair change, insertion or deletion in the nucleotide sequence of interest.

As used herein, "mutagenesis" refers to a process whereby a mutation is generated in the DNA. With "random" mutagenesis, the exact site of mutation is not predictable, occurring anywhere in the chromosome of the microorganism, and the mutation is brought about as a result of physical damage caused by agents such as radiation or
5 chemical treatment.

As used herein, "promoter" means a DNA sequence generally described as the 5' region of a gene, located proximal to the start codon. The transcription of the adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate
10 of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter.

As used herein, "percent identical" refers to the percent of the nucleotides or amino acids of the subject nucleotide or amino acid sequence that have been matched to identical nucleotides or amino acids in the compared nucleotide or amino acid sequence
15 by a sequence analysis program as exemplified below.

The present invention includes an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ
20 ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1.

It is another aspect of the present invention to provide an isolated nucleic acid molecule encoding aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide
25 consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2. Also included are proteins having SNDH activity and which are derived from a protein mentioned above by substitution, deletion, insertion or addition of one or more amino
30 acid(s) in the amino acid sequences mentioned above.

Functional derivatives as another aspect of the present invention are defined on the basis of the amino acid sequences of the present invention by addition, insertion, deletion and/or substitution of one or more amino acid residues of such sequences wherein such

derivatives still have the SNDH activity measured by an assay known in the art or specifically described herein. Such functional derivatives can be made either by chemical peptide synthesis known in the art or by recombinant techniques on the basis of the DNA sequences as disclosed herein by methods known in the state of the art. Amino acid
5 exchanges in proteins and peptides which do not generally alter the activity of such molecules are known in the state of the art.

In particular embodiments of the present invention, conservative substitutions of interest occur as follows: as example substitutions, Ala to Val/Leu/Ile, Arg to Lys/Gln/Asn, Asn to Gln/His/Lys/Arg, Asp to Glu, Cys to Ser, Gln to Asn, Glu to Asp, Gly
10 to Pro/Ala, His to Asn/Gln/Lys/Arg, Ile to Leu/Val/Met/Ala/Phe/norLeu, Lys to Arg/Gln/Asn, Met to Leu/Phe/Ile, Phe to Leu/Val/Ile/Ala/Tyr, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr/Phe, Tyr to Trp/Phe/Thr/Ser, and Val to Ile/Leu/Met/Phe/Ala/norLeu are reasonable. As preferred examples, Ala to Val, Arg to Lys, Asn to Gln, Asp to Glu, Cys to Ser, Gln to Asn, Glu to Asp, Gly to Ala, His to Arg, Ile to Leu, Leu to Ile, Lys to Arg,
15 Met to Leu, Phe to Leu, Pro to Ala, Ser to Thr, Thr to Ser, Trp to Tyr, Tyr to Phe, and Val to Leu are reasonable. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions described above, are introduced and the products screened.

Furthermore the present invention is directed to polynucleotides encoding
20 polypeptides having the SNDH activity as disclosed in the sequence listing as SEQ ID NO: 2 as well as the complementary strands, or those which include these sequences, DNA sequences or fragments thereof, and DNA sequences, which hybridize under standard conditions with such sequences but which encode for polypeptides having exactly the same amino acid sequence.

Thus, the present invention provides an isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein the complement of said nucleic acid molecule hybridizes under standard conditions with a nucleic acid molecule as described above. It is an aspect of the invention to provide an isolated nucleic acid molecule encoding a polypeptide having aldehyde dehydrogenase activity, wherein said
30 nucleic acid molecule hybridizes under standard conditions to the complementary strand of a nucleic acid molecule encoding (i) aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the nucleotide sequence of SEQ ID NO: 1; (ii) aldehyde dehydrogenase which comprises a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-
35 2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-

1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) aldehyde dehydrogenase which comprises a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609
5 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2.

"Standard conditions" for hybridization mean in this context the conditions which are generally used by a person skilled in the art to detect specific hybridization signals, or
10 preferably so called stringent hybridization conditions used by a person skilled in the art.

Thus, as used herein, the term "stringent hybridization conditions" means that hybridization will occur if there is 95% and preferably at least 97% identity between the sequences. Stringent hybridization conditions are, e.g., conditions under over night incubation at 42°C using a digoxigenin (DIG)-labeled DNA probe (constructed by using
15 a DIG labeling system; Roche Diagnostics GmbH, 68298 Mannheim, Germany) in a solution comprising 50% formamide, 5 × SSC (150 mM NaCl, 15 mM trisodium citrate), 0.2% sodium dodecyl sulfate, 0.1% N-lauroylsarcosine, and 2% blocking reagent (Roche Diagnostics GmbH), followed by washing the filters in 0.1× SSC at about 60°C.

This invention is also directed to a recombinant vector, i.e., an expression vector,
20 comprising such a nucleic acid molecule as mentioned above. An expression vector of the present invention is one which functions in a suitable host cell. Preferred vectors for the expression of the nucleic acid molecules of the present invention are vectors or derivatives thereof which are selected from the group consisting of pQE, pUC, pBluescript II, pACYC177, pACYC184, pVK100, and RSF1010.

25 A suitable host cell for expression of the nucleotide sequences of the present invention is a recombinant microorganism selected from the group consisting of bacteria, yeast, and plant cells. Preferably, the microorganism is selected from the group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia*. An example of such a preferred microorganism is *E. coli*. A more preferred
30 host cell belongs to *Gluconobacter oxydans*, most preferably *G. oxydans* DSM 4025 (FERM BP-3812), which had been deposited on March 17, 1987 under the conditions of the Budapest Treaty at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.

The microorganism "*Gluconobacter oxydans*" also includes synonyms or basonyms of such species having the same physico-chemical properties, as defined by the International Code of Nomenclature of Prokaryotes.

Thus, the present invention is directed to a recombinant microorganism which is transformed with the expression vector as described above or which comprises a nucleic acid molecule as described above integrated into its chromosomal DNA.

A wide variety of host/vector combinations may be used for cloning the double stranded nucleotide sequences of the present invention. As *E. coli* is a preferred host cell, any vectors usually used in *E. coli* are useful for the present invention. Such vectors include, but are not limited to, pQE vectors which can express His-tagged recombinant proteins (QIAGEN K.K., Tokyo, Japan), pBR322 or its derivatives including pUC18 and pBluescript II (Stratagene Cloning Systems, California, USA), pACYC177 and pACYC184 and their derivatives, and a vector derived from a broad host range plasmid such as RK2 and RSF1010. Thus, the expression vector used in the present invention is derived from pQE-plasmids, pUC-plasmids, pBluescript II, pACYC177, pACYC184, and their derivative plasmids, and a broad host range plasmid such as pVK100 and RSF1010.

As used herein, "expression vector" means a cloning vector which is capable of enhancing the expression of a gene that has been cloned into it, after transformation into a suitable host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Promoter sequences may be either constitutive or inducible.

As used herein, "cloning vector" means a plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by a single or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment may be introduced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells transformed with the cloning vector. Such markers provide, e.g., tetracycline or ampicillin resistance.

As used herein, a "recombinant vector" includes any cloning or expression vector which contains the desired cloned gene(s).

As used herein, "expression" refers to the process by which a polypeptide is produced from a structural gene. The process involves transcription of the gene into mRNA and the translation of such mRNA into polypeptide(s).

As used herein, "recombinant microorganism" includes a recombinant host which
5 may be any prokaryotic or eukaryotic cell that contains the desired cloned gene(s) on an expression or cloning vector. This term also includes those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of said microorganism.

As used herein, "host" includes any prokaryotic or eukaryotic cell that is the
10 recipient of a replicable expression vector or cloning vector. A "host", as the term is used herein, also includes prokaryotic or eukaryotic cells that can be genetically engineered by well known techniques to contain desired gene(s) on its chromosome or genome. Examples of such hosts are known to the skilled artisan.

To construct a recombinant microorganism carrying recombinant DNA, e.g., a
15 recombinant vector, various gene transfer methods including, but not limited to, transformation, transduction, conjugal mating or electroporation can be used. These methods are well-known in the field of molecular biology. Conventional transformation systems can be used for *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. A transduction system can also be used for *E. coli*. Conjugal mating
20 systems can be widely used in Gram-positive and Gram-negative bacteria including *E. coli*, *P. putida*, and *Gluconobacter*. An example of conjugal mating is disclosed in WO 89/06,688. The conjugation can occur in liquid medium or on a solid surface. Examples for a suitable recipient for SNDH production include microorganisms of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*. To the recipient for
25 conjugal mating, a selective marker may be added, e.g., resistance against nalidixic acid or rifampicin. Natural resistance can also be used, e.g., resistance against polymyxin B is useful for many *Gluconobacters*.

Preferred vectors useful for the present invention are broad-host-range vectors such as a cosmid vector like pVK100 and its derivatives and RSF1010. Copy number and
30 stability of the vector should be carefully considered for stable and efficient expression of the cloned nucleic acid molecules and also for efficient cultivation of the host cell carrying said cloned molecules. Nucleic acid molecules containing transposable elements such as Tn5 can also be used to introduce the desired DNAs into the preferred host, especially on a chromosome. Nucleic acid molecules containing any DNAs isolated

from the preferred host together with the nucleotide sequences of the present invention are also useful to introduce the nucleotide sequences of the present invention into the preferred host cell, especially on a chromosome. Such nucleic acid molecules can be transferred to the preferred host by applying any of a conventional method, e.g.,
5 transformation, transduction, conjugal mating or electroporation, which are well known in the art, considering the nature of the host cell and the nucleic acid molecule.

The nucleotide sequences including the SNDH gene provided in this invention are ligated into a suitable vector containing a regulatory region such as a promoter, a ribosomal binding site, and a transcriptional terminator operable in the host cell
10 described above with a method well-known in the art to produce a suitable expression vector.

To express the desired gene/nucleotide sequence isolated from *G. oxydans* DSM 4025 efficiently, various promoters can be used; e.g., the original promoter of the gene, promoters of antibiotic resistance genes such as kanamycin resistant gene of Tn5,
15 ampicillin resistant gene of pBR322, and beta-galactosidase of *E. coli* (lac), trp-, tac-, trc-promoter, promoters of lambda phage and any promoters which are functional in a host cell. For this purpose, the host cell can be selected from the group consisting of bacteria, yeast, and plant cells. Preferably, the host cell belongs to the genera *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* or *Escherichia*.

For expression, other regulatory elements, such as a Shine-Dalgarno (SD) sequence
20 (e.g., AGGAGG, including natural and synthetic sequences operable in the host cell) and a transcriptional terminator (inverted repeat structure including any natural and synthetic sequence operable in the host cell) which are operable in the host cell (into which the coding sequence will be introduced to provide a recombinant cell of this invention)
25 can be used with the above described promoters.

For the expression of polypeptides which are located in the periplasmic space, like the SNDH protein of the present invention, a signal peptide, which contains usually 15 to 50 amino acid residues and is totally hydrophobic, is preferably associated. A DNA encoding a signal peptide can be selected from any natural and synthetic sequence
30 operable in the desired host cell. A putative signal peptide containing amino acid residues 1-31 of SEQ ID NO: 2 was also found in the protein expressed by the SNDH gene of the present invention (SEQ ID NO: 4).

Unless otherwise mentioned, all amino acid sequences determined by sequencing the purified SNDH protein herein were determined using an automated amino acid sequencer (such as model 470A, Perkin-Elmer Applied Biosystems).

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the model ALF express II, Amersham Pharmacia Biotech), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of the DNA sequence determined as above. Therefore, as it is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The invention provides an isolated nucleic acid molecule encoding the enzyme (SNDH). Methods and techniques designed for the manipulation of isolated nucleic acid molecules are well known in the art. Methods for the isolation, purification, and cloning of nucleic acid molecules, as well as methods and techniques describing the use of eukaryotic and prokaryotic host cells and nucleic acid and protein expression therein, are known to the skilled person.

Briefly, the SNDH gene, the DNA molecule containing said gene, the recombinant expression vector and the recombinant microorganism used in the present invention can be obtained by the following steps:

- (1) isolating chromosomal DNA from *G. oxydans* DSM 4025 and constructing a gene library with the chromosomal DNA in an appropriate host cell, e.g., *E. coli*;
- (2) cloning the SNDH gene from the chromosomal DNA by colony-, plaque-, or Southern-hybridization, PCR (polymerase chain reaction) cloning, western-blot analysis or other techniques known in the art;

- (3) determining the nucleotide sequence of the SNDH gene obtained as above by conventional methods to select a DNA molecule containing said SNDH gene and constructing a recombinant expression vector on which the SNDH gene can be expressed efficiently;
- 5 (4) constructing recombinant microorganisms carrying the SNDH gene by an appropriate method for introducing DNA into a host cell, e.g., transformation, transduction, conjugal transfer and/or electroporation, which host cell thereby becomes a recombinant microorganism of this invention.

The materials and techniques used in the above aspect of the present invention are
10 exemplified in detail as follows:

A total chromosomal DNA can be purified by a procedure well known in the art. The desired gene can be cloned in either plasmid or phage vectors from a total chromosomal DNA typically by either of the following illustrative methods:

- (i) The partial amino acid sequences are determined from the purified proteins or
15 peptide fragments thereof. Such whole protein or peptide fragments can be prepared by the isolation of such a whole protein or by peptidase-treatment from the gel after SDS-polyacrylamide gel electrophoresis. Thus obtained protein or fragments thereof are applied to protein sequencer such as Applied Biosystems automatic gas-phase sequencer 470A. The amino acid sequences can be utilized to design and prepare oligonucleotide
20 probes and/or primers with DNA synthesizer such as Applied Biosystems automatic DNA sequencer 381A. The probes can be used for isolating clones carrying the target gene from a gene library of the strain carrying the target gene by means of Southern-, colony- or plaque-hybridization.
- (ii) Alternatively, for the purpose of selecting clones expressing a target protein from the
25 gene library, immunological methods with antibodies prepared against the target protein can be applied.
- (iii) The DNA fragment of the target gene can be amplified from the total chromosomal DNA by PCR method with a set of primers, i.e., two oligonucleotides synthesized according to the amino acid sequences determined as above. Then a clone carrying the
30 target-whole gene can be isolated from the gene library constructed, e.g., in *E. coli*, by Southern-, colony-, or plaque-hybridization with the PCR product obtained above as probe.

DNA sequences which can be made by PCR by using primers designed on the basis of the DNA sequences disclosed herein by methods known in the art are also an object of
35 the present invention.

Above mentioned antibodies can be prepared with the purified SNDH proteins, the purified recombinant SNDH proteins such as His-tagged SNDH expressed in *E. coli*, or its peptide fragment as an antigen.

Once a clone carrying the desired gene is obtained, the nucleotide sequence of the target gene can be determined by a well known method such as dideoxy chain termination method with M13 phage.

Brief description of the Figures:

Figure 1 illustrates the gene encoding a protein having aldehyde dehydrogenase activity of the present invention. The restriction map of the SNDH gene and ORF-A gene is given, wherein ORF means open reading frame and Signal seq. means the putative signal peptide sequence of the SNDH gene.

Figure 2 illustrates the restriction map of the SNDH and ORF-A genes cloned into cosmid pVSN5 as well as cloning of insert DNA of different sizes into pUC plasmids pUCSNP4, pUCSNP9, pUCSN19, and pUCSN5. In the physical map of pVSN5, the arrow filled in gray shows the SNDH gene.

Figure 3 shows the cloning strategy of the 8.0 kb *Pst*I fragment including the intact SNDH gene from pVSN5 into pUC18 vectors to result in pUCSNP4 and pUCSNP9. Note that pUCSNP4 and pUCSNP9 are identical except for the direction of the insert.

Figure 4 shows schematically the construction of the GOMTR1SN::Km (SNDH-disruptant) by using a suicide vector plasmid having the disrupted SNDH gene with kanamycin cassette (Km). Homologous recombination between the vector plasmid and the chromosomal DNA of GOMTR1 as the parent strain at the corresponding region occurs to obtain the disruptant strains. "G.O." means *Gluconobacter oxydans*.

Figure 5 represents the physical map of the insert DNA of pVSN117, pVSN106, and pVSN114. Plasmid pVSN117 has the insert DNA encoding the C-terminal deleted SNDH gene (nucleotides 258-1955 of SEQ ID NO: 1, i.e., amino acids 1-566 of SEQ ID NO: 2), which expresses only a 55 kDa protein. Plasmids pVSN106 and pVSN114 have the insert DNA encoding the intact SNDH gene.

The gene of the present invention encodes an SNDH enzyme of 578 amino acid residues (SEQ ID NO: 5 consisting of amino acids 32-609 of SEQ ID NO: 2) together with a putative signal peptide of 31 amino acid residues (SEQ ID NO: 4 consisting of

amino acids 1-31 of SEQ ID NO: 2) as depicted in Figure 2. In terms of nucleotide sequences, the coding region of the SNDH gene encompasses nucleotides 258-2087 of SEQ ID NO: 1 and includes the coding sequences for a putative signal peptide (nucleotides 258-350 of SEQ ID NO: 1) and the stop codon (nucleotides 2085-2087 of
5 SEQ ID NO: 1). Thus, the nucleotide sequence without the stop codon is the nucleotide sequence from position 258-2084 of SEQ ID NO: 1, and additionally without the signal sequence the nucleotide sequence encompasses nucleotides 351-2084 of SEQ ID NO: 1.

The nucleic acid molecules as disclosed in the present invention can be used in a recombinant microorganism for the production of 2-KGA and/or vitamin C. The
10 recombinant microorganism, which is selected from *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia*, may be cultured in an aqueous medium supplemented with appropriate nutrients under aerobic conditions. The cultivation may be conducted at a pH of 4.0 to 9.0, preferably 6.0 to 8.0. The cultivation period varies depending on the pH, temperature and nutrient medium to be used, and is
15 preferably about 1 to 5 days. The preferred temperature range for carrying out the cultivation is from about 13°C to about 36°C, preferably from about 18°C to about 33°C. It is usually required that the culture medium contains such nutrients as assimilable carbon sources, e.g., glycerol, D-mannitol, D-sorbitol, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, D-ribose, D-fructose, D-glucose, and sucrose, preferably D-sorbitol, D-
20 mannitol, and glycerol; and digestible nitrogen sources such as organic substances, e.g., peptone, yeast extract, baker's yeast, urea, amino acids, and corn steep liquor. Various inorganic substances may also be used as nitrogen sources, e.g., nitrates and ammonium salts. Furthermore, the culture medium usually contains inorganic salts, e.g., magnesium sulfate, potassium phosphate, and calcium carbonate. The cultivation is carried out in
25 appropriate equipment such as jar fermentors, flasks, or tubes. The recombinant microorganism is either transformed with an expression vector containing the nucleic acid molecules as above or comprises the nucleic acid molecule of the present invention integrated into its chromosomal DNA.

An appropriate sugar compound used as substrate for the production of 2-KGA
30 and/or vitamin C is L-sorbose. The metabolic pathway for 2-KGA and vitamin C goes from D-sorbitol via L-sorbose to L-sorbose, which is then converted to 2-KGA and/or vitamin C. Thus, the direct substrate for both products is L-sorbose.

Thus, it is an aspect of the present invention to provide a process for the production of 2-KGA and/or vitamin C from L-sorbose comprising (a) propagating or
35 cultivating the recombinant microorganism, which is either transformed with an

the present invention or which comprises the nucleic acid molecules of the present invention integrated into its chromosomal DNA, in an appropriate culture medium and (b) recovering and separating 2-KGA and/or vitamin C from said culture medium.

It is one embodiment to provide a process for the production of vitamin C and/or 2-KGA from L-sorbose comprising (a) propagating a recombinant organism in an appropriate culture medium, wherein the nucleic acid molecule as of the present invention is heterologously introduced to said recombinant organism, and (b) recovering and separating vitamin C and/or 2-KGA from said culture medium.

The present invention provides recombinant SNDH. One can increase the production yield of the SNDH enzyme by introducing the SNDH gene provided by the present invention into a host cell including *G. oxydans* DSM 4025. One can also produce more efficiently the SNDH proteins in a host cell selected from a group consisting of *Gluconobacter*, *Acetobacter*, *Pseudomonas*, *Acinetobacter*, *Klebsiella* and *Escherichia* by using the SNDH gene of the present invention. The microorganism may be cultured as described above.

An embodiment for the isolation and purification of the recombinant SNDH from the microorganism after the cultivation is briefly described hereinafter: cells are harvested from the liquid culture broth by centrifugation or filtration. The harvested cells are washed with water, physiological saline or a buffer solution having an appropriate pH. The washed cells are suspended in the buffer solution and disrupted by means of a homogenizer, sonicator or French press, or by treatment with lysozyme to give a solution of disrupted cells. The recombinant SNDH is isolated and purified from the cell-free extract of disrupted cells, preferably from the cytosol fraction of the microorganism. The recombinant SNDH can be immobilized on a solid carrier for solid phase enzyme reaction.

The invention is further directed to a process for the production of 2-KGA from L-sorbose comprising (a) cultivating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture medium, wherein the gene encoding aldehyde dehydrogenase represented by SEQ ID NO: 2 is disrupted in said microorganism, and (b) recovering and separating 2-KGA from said culture medium. The disruption may take place anywhere in the gene, resulting in a non-functioning of the encoded enzyme.

Thus, a process is provided for the production of 2-KGA via L-sorbose from an appropriate sugar compound comprising (a) propagating a microorganism belonging to *Gluconobacter oxydans* DSM 4025 in an appropriate culture medium, wherein the gene encoding aldehyde dehydrogenase is disrupted in said microorganism, said aldehyde dehydrogenase being encoded by (i) a polynucleotide being at least 95% identical to the nucleotide sequence of SEQ ID NO: 1; (ii) a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) a polynucleotide selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2. The resulting 2-KGA is further recovered and isolated from said culture medium.

In one embodiment, the invention provides a process for the disruption of the SNDH gene by classical mutagenesis with agents such as UV-irradiation or chemical treatment by any mutation reagents, e.g., *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG), ICR170 or acrydine orange, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for the disruption of the SNDH gene by DNA recombination techniques such as transposon insertion or site directed mutagenesis by PCR, *in vivo* as well as *in vitro*.

In another embodiment, the invention provides a process for producing 2-KGA using the disruptants described above by fermentation from an appropriate substrate, i.e., a sugar compound, which is selected from the group consisting of L-sorbose, D-glucose, D-sorbitol, and L-sorbose. The process takes place in appropriate equipment such as jar fermentors, flasks, or tubes. Furthermore, the invention provides a process for producing 2-KGA using a cell free extract of the disruptants described above by incubation from an appropriate substrate, e.g., L-sorbose, D-glucose, L-sorbose, and D-sorbitol, in appropriate equipment such as a bioreactor.

The present invention provides recombinant SNDH. Furthermore, it is directed to a process for the production of aldehyde dehydrogenase comprising (a) cultivating a recombinant microorganism comprising a nucleic acid molecule encoding aldehyde dehydrogenase which comprises (i) a polynucleotide being at least 95% identical to the

nucleotide sequence of SEQ ID NO: 1; (ii) a polynucleotide being at least 95% identical to the polynucleotide selected from the group consisting of (a) nucleotides 258-2084 of SEQ ID NO: 1, (b) nucleotides 351-2084 of SEQ ID NO: 1, (c) nucleotides 258-1955 of SEQ ID NO: 1, and (d) nucleotides 351-1955 of SEQ ID NO: 1; and (iii) a polynucleotide
5 selected from the group consisting of (a) a polynucleotide encoding the polypeptide having the amino acid sequence of SEQ ID NO: 2, (b) a polynucleotide encoding the polypeptide consisting of amino acids 32-609 of SEQ ID NO: 2, (c) a polynucleotide encoding the polypeptide consisting of amino acids 1-566 of SEQ ID NO: 2, and (d) a
10 polynucleotide encoding the polypeptide consisting of amino acids 32-566 of SEQ ID NO: 2; wherein said microorganism is cultivated in an appropriate culture medium and (b) wherein said aldehyde dehydrogenase is recovered and separated from said culture medium.

Example 1: Amino acid sequencing from the N-terminus of SNDH

15 The partial amino acid sequence of the N-terminal 75 kDa subunit of the SNDH protein was determined. About 10 µg of the SDS-treated purified SNDH consisting of 75 kDa subunits was subjected to SDS-PAGE, and the protein band was electroblotted onto a PVDF membrane. The protein blotted on the membrane was soaked in a digestion buffer (100 mM potassium phosphate buffer, 5 mM dithiothreitol, 10 mM EDTA, pH
20 8.0) and incubated with 5.04 µg of pyroglutamate aminopeptidase (SIGMA, USA) at 30°C for 24 hours. After incubation, the membrane was washed with deionized water and subjected to N-terminal amino acid sequencing using an automated amino acid sequencer (ABI model 490, Perkin Elmer Corp., Conn., USA). As a result, 14 residues of the N-terminal amino acid sequence were obtained as illustrated in SEQ ID NO: 3.

25

Example 2: Cloning of partial SNDH gene by PCR

Amplification of the partial SNDH gene fragment was carried out by PCR with chromosomal DNA of *G. oxydans* DSM 4025 (FERM BP-3812) and degenerated oligonucleotide DNA primers, P11 (SEQ ID NO: 6) and P12 (SEQ ID NO: 7). Both
30 primers were degenerated DNA mixtures having bias for *Gluconobacter* codon usage. The PCR was performed with thermostable taq polymerase (TAKARA Ex Taq™, Takara Shuzo Co., Ltd., Seta 3-4-1, Otsu, Shiga, 520-2193, Japan), using a thermal cycler (Gene

Amp PCR System 2400-R, PE Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA). The reaction mixture (25 µl) consisted of 200 µM of dNTPs, 50 pmol of each primer (24 ~ 48 degeneracy), 5 ng of the chromosomal DNA, and 1.25 units of the DNA polymerase in the buffer provided from the supplier. The reaction was carried out with 5 cycles of denaturation step at 94°C for 30 sec, annealing step at 37°C for 30 sec, synthesis step at 70°C for 1 min followed by 25 cycles of denaturation step at 94°C for 30 sec, annealing step at 50°C for 30 sec, synthesis step at 70°C for 1 min. As a result, a 41 bp DNA fragment was specifically amplified and cloned into vector pCR 2.1-TOPO (Invitrogen, 1600 Faraday Avenue Carlsbad, California 92008, USA) to obtain a recombinant plasmid pMTSN2. The nucleotide sequence of the cloned 41 bp DNA fragment which encodes an N-terminal partial amino acid sequence of the mature SNDH protein was confirmed by dideoxy-chain termination method (F. Sanger et al, Proc. Natl. Acad. Sci. USA, 74, 5463-5467, 1977).

Example 3: Complete cloning of the SNDH gene

(1) Construction of a gene library of *G. oxydans* DSM 4025

The chromosomal DNA of *G. oxydans* DSM 4025 was prepared from cells grown on M agar medium containing 5% D-mannitol, 1.75% corn steep liquor, 5% baker's yeast, 0.25% MgSO₄·7H₂O, 0.5% CaCO₃ (Practical Grade), 0.5% urea, and 2.0% agar (pH 7.0), for 4 days at 27°C. The chromosomal DNA (4 µg) was partially digested with 4 units of *Eco*R I in 20 µl of reaction mixture. A portion (8 µl) of the sample containing partially-digested DNA fragments was separated by an electrophoresis using 1% agarose gel. Fragments ranging from 15 to 35 kb were cut out and chemically melted to recover the fragments using QIAEX II (QIAGEN Inc., 28159 Avenue Stanford, Valencia, CA 91355, USA). The objective DNA fragments recovered were suspended in H₂O. On the other hand, 2 µg of a cosmid vector pVK100 was completely digested with *Eco*R I and dephosphorylated of the 5'-ends by treating with bacterial alkaline phosphatase (*E. coli* C75) (Takara Shuzo). The treated pVK100 (220 ng) was ligated with the 15 - 35 kb *Eco*R I fragments (1 µg) using a ligation kit (Takara Shuzo) in 36 µl of reaction mixture. The ligated DNA which had been ethanol precipitated and resolved in appropriate volume of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) was used for *in vitro* packaging (Gigapack III Gold Packaging Extract, Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA) to infect *E. coli* VCS257, a host strain for the genomic library. As a

result, totally 400,000 - 670,000 clones containing about 25 kb-inserted DNA fragments were obtained.

(2) Complete cloning of the SNDH gene by colony hybridization

The probe that would be used for screening of the cosmid library described above to detect clones carrying the complete SNDH gene by colony hybridization method, was constructed. The 41 bp DNA fragment encoding the N-terminal amino acid sequence of SNDH was amplified and labeled by PCR-DIG labeling method (Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alabama, CA94501, USA). PCR with plasmid pMTSN2 DNA as a template and oligonucleotide DNA primers, P13 (SEQ ID NO: 8) and P14 (SEQ ID NO: 9), was performed with thermostable taq polymerase (TAKARA Ex TaqTM, Takara Shuzo Co., Ltd.), using a thermal cycler (Gene Amp PCR System 2400-R, PE Biosystems). The reaction was carried out with 25 cycles of denaturation step at 94°C for 30 sec, annealing step at 55°C for 30 sec, synthesis step at 70°C for 1 min. Using the DIG-labeled probe, screening of the cosmid library (about 1,000 clones) by colony hybridization and chemiluminescent detection according to the method provided from the supplier (Roche Molecular Systems Inc., USA) was carried out. As a result, 3 positive clones were isolated and one of them was designated pVSN5, which carried about 25 kb insert DNA in pVK100 vector. From this, 25 kb DNA insert fragments of different sizes were further subcloned into pUC18 vectors (Figure 2): (1) a 3.2 kb *EcoR* I fragment comprising the upstream portion (the N-terminal part) of the SNDH gene resulting in pUCSN19, (2) a 7.2 kb *EcoR* I fragment comprising the downstream portion (the C-terminal part) of the SNDH gene resulting in pUCSN5, and (3) a 1.8 kb *Pst* I fragment comprising the intact or complete SNDH gene resulting in pUCSNP4 and pUCSNP9, respectively. Note that the inserts in pUSNP4 and pUSNP9 are the same but in opposite directions.

(3) Nucleotide sequencing of the SNDH gene

Plasmids pUCSN19, pUCSN5, and pUCSNP4 were used for nucleotide sequencing of a region including the SNDH gene or gene fragments. The determined nucleotide sequence (SEQ ID NO: 1 with 3,408 bp) revealed that the ORF of the SNDH gene (1,827 bp; nucleotides 258-2084 of SEQ ID NO: 1) encoded a polypeptide of 609 amino acid residues (SEQ ID NO: 2). An additional ORF, ORF-A, was found downstream of the SNDH ORF as illustrated in Figure 1. ORF-A (1,101 bp; nucleotides 2214-3314 of SEQ ID NO: 1) encoded a polypeptide of 367 amino acids.

In the ORF of the SNDH gene, a signal peptide-like sequence (SEQ ID NO: 4 with 31 amino acids) is possibly included in the deduced amino acid sequence, which contains (i) many hydrophobic residues, (ii) positively-charged residues close to the N-terminus and (iii) Ala-Xaa-Ala motif for cleavage site of the signal sequence. The putative
5 ribosome-binding site (Shine-Dalgarno, SD, sequence) for the SNDH gene was located at 6 bp upstream of the initiation codon (AGGAGA at nucleotide position 247-252 of SEQ ID NO: 1). Furthermore, a motif (Cys-Xaa-Xaa-Cys-His) defined as heme c binding site was found at position 530-534 of SEQ ID NO: 2. From the genetically analysis as shown above, the SNDH protein is thought to be one of quinoxemoproteins.

10 A homology search for the SNDH gene using the program of FASTA in GCG (Genetics Computer Group, Madison, WI, USA) revealed that Arg227, Asn228, Gln230, Gly246, and Asp251 of SEQ ID NO: 5 correspond to several highly conserved residues in the presumed active site of *A. calcoaceticus* GDH-B protein described by Oubrie et al. [J. Mol. Biol. 289:319-333 (1999)].

15

Example 4: Expression of the SNDH gene in *E. coli*

Plasmids pUCSNP4 and pUCSNP9 (Figure 3), containing the 8.0 kb *Pst* I-fragment with the intact, i.e., complete SNDH gene, were transformed into *E. coli* JM109 to confirm the expression and the activity of the SNDH proteins. The amount of vitamin C
20 produced as the enzyme activity was measured at a wavelength of 264nm by a high performance liquid chromatography system (HPLC) which was composed with a UV detector (TOSOH UV8000; Tosoh, Japan), a dualpump (TOSOH CCPE; Tosoh), an integrator (Shimadzu C-R6A; Shimadzu, Japan) and a column (YMC-Pack Polyamine-II; 4.6 mm of inner diameter [i.d.] × 15 cm, YMC, U.S.A.).

25 The conversion activity of L-sorbose to vitamin C by using cytosol fraction of the recombinant *E. coli* was tested (Table 1). Cells were cultivated in LB medium optionally supplemented with 10 μM of PQQ and 1.0 mM of CaCl₂. The cytosol fraction was prepared by ultracentrifugation (100,000 × g, 45 min) of the cell free extract in 50 mM potassium phosphate buffer (pH 7.0). The reaction mixture (100 μl) consisted of
30 125 μg of cytosol fraction of the recombinant *E. coli*, 50 mM of L-sorbose, 1.0 mM of phenazine mesosulfate (PMS), with or without the addition of 1.0 μM of PQQ and 1.0 mM of CaCl₂ as cofactors, depending on the case. The enzyme reaction was carried out at 30°C for 30 min. The holo-SNDHs of the cells cultivated in LB medium containing 10

μM of PQQ and 1.0 mM of CaCl_2 produced vitamin C definitely under the defined reaction condition without the cofactors of PQQ and CaCl_2 . By addition of the cofactors, the apo-SNDH expressed with pUCSNP4 and pUCSNP9 showed almost the same activity as those of the holo-enzyme.

5

Table 1: activity measurement of recombinant SNDH

Microorganism	PQQ and CaCl_2 in the medium	Specific activity (mU/mg Protein)	
		with PQQ and CaCl_2	without PQQ and CaCl_2
<i>E. coli</i> JM109/pUCSNP4	+	0.187	0.224
<i>E. coli</i> JM109/pUCSNP9	+	0.198	0.252
<i>E. coli</i> JM109/pUC18	+	0.000	0.000
<i>E. coli</i> JM109/pUCSNP4	-	0.155	0.000
<i>E. coli</i> JM109/pUCSNP9	-	0.176	0.000
<i>E. coli</i> JM109/pUC18	-	0.000	0.000
<i>G. oxydans</i> DSM 4025	-	0.026	0.026

One unit (U) of the enzyme was defined as the amount of enzyme, which produces 1.0 mg of vitamin C in the defined reaction.

10 Example 5: Construction and cultivation of SNDH-gene disruptants of *G. oxydans* strains

Figure 4 shows the scheme for the construction of SNDH gene targeting vector, GOMTR1SN::Km (SNDH-disruptant). First, plasmid pSUPSN was constructed by a ligation of the 8.0 kb *Pst* I fragment containing the SNDH gene from plasmid pUCSNP4 with a suicide vector pSUP202 (for reference see Simon et al., A broad host range mobilization system for in vitro genetic engineering: transposon mutagenesis in Gram negative bacteria, Biotechnology, 1, 784-791, 1983). Second, a kanamycin-resistant-gene cassette (Km cassette) was inserted into the *Eco*R I site of the SNDH gene cloned in plasmid pSUPSN to obtain plasmid pSUPSN::Km (Km^rTc^r). Then, plasmid pSUPSN::Km was introduced into GOMTR1, which was a rifampicin (Rif) resistant derived spontaneously from wildtype *G. oxydans* DSM 4025 strain, to result in SNDH-null mutants ($\text{Km}^r\text{Rif}^r\text{Tc}^r$).

GOMTR1 was cultivated in a 200 ml flask containing 50 ml of T broth, which was composed of 30 g/l of Trypticase Soy Broth (BBL; Becton Dickinson and Company, Cockeysville, MD 21030, USA) and 3 g/l of yeast extract (Difco; Becton Dickinson Microbiology Systems, Becton Dickinson and Company, Sparks, MD 21152, USA) with
5 100 µg/ml of rifampicin at 30°C overnight. *E. coli* HB101 (pRK2013) [D. H. Figurski, Proc. Natl. Acad. Sci. USA, 76, 1648-1652, 1979] and *E. coli* JM109 (pSUPSN::Km) were cultivated in test tubes containing 2 ml of LB medium with 50 µg/ml of kanamycin at 30°C overnight. Cultured cells of GOMTR1, *E. coli* HB101 (pRK2013), and *E. coli* JM109 (pSUPSN::Km) were collected separately by centrifugation and each cell suspension in LB
10 medium was mixed in the ratio of 10:1:1, respectively. Then these cell suspensions were mixed at the same volume and the mixture was spread out on a 0.45 µm nitrocellulose membrane (PROTRAN, Schleicher & Schuell GmbH, Postfach 4, D-37582 Dassel, Germany) put on an agar medium, which was composed of 5.0% mannitol, 0.25% MgSO₄·7H₂O, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% CaCO₃, and
15 2.0% agar, for conjugal transfer of the suicide plasmid from the *E. coli* donor to GOMTR1 as recipient. After cultivation at 27°C for 1 day, the cells containing transconjugants were suspended and diluted appropriately with T broth, and spread out on the screening agar plates containing 100 µg/ml of rifampicin and 50 µg/ml of kanamycin. Finally, several transconjugants (Km^rRif^rTc^s) which had the disrupted
20 SNDH gene with Km cassette were obtained.

GOMTR1 and the disruptants, GOMTR1SN::Km, were grown on an agar plate containing 8.0% L-sorbose, 0.25% MgSO₄·7H₂O, 1.75% corn steep liquor, 5.0% baker's yeast, 0.5% urea, 0.5% CaCO₃; and 2.0% agar at 27°C for 4 days. One loopful of the cells was inoculated into 50 ml of a seed culture medium (pH 6.0) containing 4% D-sorbitol,
25 0.4% yeast extract, 0.05% glycerol, 0.25% MgSO₄·7H₂O, 1.75% corn steep liquor, 0.1% urea, and 1.5% CaCO₃ in a 500 ml Erlenmeyer flask, and cultivated at 30°C with 180 rpm for 1 day on a rotary shaker. The seed culture thus prepared was used for inoculating 50 ml of a main culture medium, which was composed of 12.0% L-sorbose, 2.0% urea, 0.05% glycerol, 0.25% MgSO₄·7H₂O, 3.0% corn steep liquor, 0.4% yeast extract, and
30 1.5% CaCO₃ in a 500 ml Erlenmeyer flask. The cultivation was carried out at 30°C and 180 rpm for 4 days. The activity assay was performed as described in Example 4. The amount of 2-KGA produced as the enzyme activity was measured at a wavelength of 340nm by a HPLC system that was composed with a UV detector (TOSOH UV8000; Tosoh), a dualpump (TOSOH CCPE; Tosoh), an integrator (Shimadzu C-R6A; Shimadzu) and a column (YMC-Pack Pro C18, YMC). As shown in Table 2, the
35

production efficiency for 2-KGA of the SNDH-gene disruptants was higher than that of the parent strain GOMTR1. The difference of the conversion rate per mol L-sorbose to 2-KGA was about 3%.

5 Table 2: 2-KGA production of *G. oxydans* strains having a disrupted SNDH-gene

Strain	2-KGA (g/L)	Residual L-sorbose (g/L)	*Molar yield (mol %)
GOMTR1SN::Km	96.7	15.3	99.2
GOMTR1	98.8	9.8	95.5

*Molar yield: mol 2-KGA produced/mol L-sorbose consumed.

Example 6: Introduction of the plasmids carrying the SNDH gene into the SNDH-gene disruptant of *G. oxydans* DSM 4025

10 Several kinds of SNDH-expression plasmids using broad host range vector pVK100 were constructed as shown in Figure 5. Those plasmids have different insert DNAs at the *Hind* III site of pVK100 described as follows: pVSN117 has the insert DNA containing the incomplete SNDH gene encoding a polypeptide ending at Gly535 of SEQ ID NO: 5 (amino acid residue 566 of SEQ ID NO: 2), i.e., a C-terminal deleted SNDH gene, which
 15 expresses only a 55 kDa protein. Plasmids pVSN106 and pVSN114, respectively, have the insert DNA containing the complete SNDH gene. Those plasmids were introduced into strain GOMTR1SN::Km by conjugal transfer method.

The transconjugants having the plasmids shown in Figure 5 were grown on an agar plate containing 10.0% L-sorbose, 0.25% MgSO₄·7H₂O, 1.75% corn steep liquor, 5.0%
 20 baker's yeast, 0.5% urea, 0.5% CaCO₃, and 2.0% agar at 27°C for 4 days. The enzyme reaction mixture consisted of 80 µg of cell free extract of the recombinant *Gluconobacter* strains, 25 mM potassium phosphate buffer (pH 7.0), 50 mM of L-sorbose, and 0.05 mM of PMS. The enzyme reaction was carried out at 30°C for 30 min with shaking at 1,000 rpm. The activity assay was performed according to Example 4. The result is
 25 shown in Table 3.

Table 3: production of vitamin C with different SNDH constructs

Host cell	Vector DNA	Vitamin C produced (mg/L)
GOMTR1SN::Km-2	pVK100	0.0
	pVSN117	473.2
	pVSN106	845.3
	pVSN114	860.2